Simulating the Bootstrap Rejection Probability

1. The Bootstrap Discrepancy

A single bootstrap test is to be based on a statistic τ in asymptotic P value form. Rejection by an asymptotic test at level α is thus the event $\tau < \alpha$. Rejection by the bootstrap test is the event $\tau < Q(\alpha, \mu^*)$, where μ^* is the bootstrap DGP, and $Q(\alpha, \mu^*)$ is the (random) $\alpha-$quantile of the distribution of the statistic τ as generated by μ^*.

We define two random variables that are deterministic functions of the two random elements, τ and μ^*, needed for computing the bootstrap P value $R(\tau, \mu^*)$, where R is the inverse function of Q, in the sense that, for all $\alpha \in [0, 1]$, and for all DGPs μ in the model we consider,

$$\alpha = R(Q(\alpha, \mu), \mu) = Q(R(\alpha, \mu), \mu). \tag{1}$$

The first of these random variables is distributed as $\mathrm{U}(0, 1)$ under μ; it is

$$p \equiv R(\tau, \mu). \tag{2}$$

The uniform distribution of p follows from the fact that $R(\cdot, \mu)$ is the CDF of τ under μ. The second random variable is

$$r \equiv R(Q(\alpha, \mu^*), \mu). \tag{3}$$

We may rewrite the event which leads to rejection by the bootstrap test at level α as $R(\tau, \mu) < R(Q(\alpha, \mu^*), \mu)$, by acting on both sides of the inequality $\tau < Q(\alpha, \mu^*)$ by the increasing function $R(\cdot, \mu)$. With the definitions (2) and (3), this event becomes simply $p < r$. Let the CDF of r under μ conditional on the random variable p be denoted as $F(r \,|\, p)$. Then the probability under μ of rejection by the bootstrap test at level α is

$$\begin{aligned} \mathrm{E}(\mathrm{I}(p < r)) &= \mathrm{E}\big(\mathrm{E}(\mathrm{I}(p < r) \,|\, p)\big) = \mathrm{E}\big(\mathrm{E}(\mathrm{I}(r > p) \,|\, p)\big) \\ &= \mathrm{E}\big(1 - F(p \,|\, p)\big) = 1 - \int_0^1 F(p \,|\, p)\, dp, \end{aligned} \tag{4}$$

since the marginal distribution of p is $\mathrm{U}(0, 1)$.

The bootstrap discrepancy is defined as the difference between the rejection probability (RP) of the bootstrap test and the nominal significance level α. A useful expression for the bootstrap discrepancy is obtained by defining the random variable $q \equiv r - \alpha$. The CDF of q conditional on p is then $F(\alpha + q \,|\, p) \equiv G(q \,|\, p)$. The RP (4) minus α is

$$1 - \alpha - \int_0^1 G(p - \alpha \,|\, p)\, dp.$$

Changing the integration variable from p to $x = p - \alpha$ gives for the bootstrap discrepancy

$$\begin{aligned} & 1 - \alpha - \int_{-\alpha}^{1-\alpha} G(x \,|\, \alpha + x)\, dx \\ &= 1 - \alpha - \Big[x\, G(x \,|\, \alpha + x) \Big]_{-\alpha}^{1-\alpha} + \int_{-\alpha}^{1-\alpha} x\, dG(x \,|\, \alpha + x) \\ &= \int_{-\alpha}^{1-\alpha} x\, dG(x \,|\, \alpha + x), \end{aligned} \tag{5}$$

because $G(-\alpha \,|\, 0) = F(0 \,|\, 0) = 0$ and $G(1 - \alpha \,|\, 1) = F(1 \,|\, 1) = 1$.

To a very high degree of approximation (at least in all cases considered so far), (5) can be replaced by

$$\int_{-\infty}^{\infty} x\, dG(x \,|\, \alpha), \tag{6}$$

that is, the expectation of q conditional on p being at the margin of rejection at level α. In cases in which p and q are independent or nearly so, it may be a good approximation to replace (6) by the unconditional expectation of q.

2. Basic Setup for Simulation

We now wish to see how expressions (5) and (6), along with the unconditional expectation of q, can be estimated by simulation. To that end, it is convenient to define some more explicit notation. We start with a probability space $(\Omega, \mathcal{F}, P)$, where each point $\boldsymbol{\omega} \in \Omega$ contains enough information to generate a sample of size n for any given DGP. Note that the sample size is held fixed throughout the present analysis.

The DGPs of most models are specified in terms of some deterministic (or exogenous) information, and some probability distributions that are used to generate realisations of random elements. We can safely assume that the exogenous information is fixed once and for all in all that follows. Random elements are realised as deterministic functions, specific to a particular DGP, of a set of m IID drawings from the U$(0, 1)$ distribution. The number m is often equal to the sample size n if the model has only one dependent variable, but, for models with several dependent variables, m can be an integer multiple of n. If we can work with sufficient statistics, m may well be small and independent of the sample size. We then set $\Omega = (0, 1)^m$, and P becomes Lebesgue measure on Ω, that is, the measure of the uniform distribution on the unit hypercube. A drawing of a point $\boldsymbol{\omega}$ is to be thought of as the *set of random numbers* used for a particular simulation.

As an example, suppose that the DGP for two dependent variables, $\boldsymbol{x}$ and $\boldsymbol{y}$, can be written as

$$\begin{gathered} y_t = \alpha_0 + x_t \beta_0 + u_t, \quad x_t = \gamma_0 + \delta_0 w_t + v_t, \quad t = 1, \ldots, n, \\ \begin{bmatrix} u_t \\ v_t \end{bmatrix} \sim \text{NID} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho \sigma_u \sigma_v \\ \rho \sigma_u \sigma_v & \sigma_v^2 \end{bmatrix} \right), \end{gathered} \tag{7}$$

where α_0, β_0, γ_0, and δ_0 are fixed parameters, and the w_t are given exogenously. For this DGP, we have $\Omega = (0,1)^{2n}$, and in order to construct a realised sample, we begin by drawing a point $\boldsymbol{\omega} \in \Omega$ from the uniform distribution:

$$\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_{2n}).$$

We next generate a vector $\boldsymbol{z}$ of $2n$ independent standard normal variables, with typical element $z_s = \Phi^{-1}(\omega_s)$. For each $t = 1, \ldots, n$, we can then set

$$u_t = \sigma_u z_t, \quad v_t = \rho\sigma_v u_t/\sigma_u + \sigma_v(1-\rho^2)^{1/2} z_{n+t},$$

so as to generate the (u_t, v_t) pairs according to the specification of the DGP. The actual dependent variables (y_t, x_t) are then computed simply using (7).

This DGP implicitly makes use of two mappings from Ω to $\mathbb{R}$, one to generate the x_t, the other to generate the y_t. If we wish to consider a test statistic τ, it can be obtained by composing the mappings for the x_t and the y_t with the mapping that computes the statistic as a function of given data.

For bootstrapping, the first essential is to set up a bootstrap DGP. This DGP, like any other DGP, depends on parameters and probability distributions, and its randomness comes from the fact that these constitutive elements are random. Like test statistics, realisations of parameter estimates are obtained by deterministic, DGP-specific, mappings from Ω to a parameter space. For parametric bootstraps, that is enough. For non- or semi-parametric bootstrap, we usually need a set of objects to resample. These too are functions of the data, and so are generated by deterministic DGP-specific mappings from Ω.

The usual sort of setup is thus as follows. A vector-valued mapping that is specific to a given DGP defines realisations of samples, or of a set of sufficient statistics, in terms of drawings from the uniform distribution on Ω. We can denote such a mapping as $\boldsymbol{y}(\mu, \boldsymbol{\omega})$, where we make explicit the fact that this transformation from uniform data depends on the DGP μ. Next, we have mappings, that do not depend on the DGP, that transform data into objects of interest, test statistics, parameter estimates, bootstrap DGPs, *etc.* For present purposes, we do not distinguish the first, DGP-specific, transformations from the second, non-specific, ones. We write simply $\tau(\mu, \boldsymbol{\omega})$ for the realisation of the statistic of interest from the DGP μ obtained from a uniform realisation $\boldsymbol{\omega} \in \Omega$, and we similarly write $b(\mu, \boldsymbol{\omega})$ for a realisation of the bootstrap DGP from the same realised data. The sort of bootstrap DGP that is desired is thus embedded in the mapping b.

3. Bootstrap Samples

Bootstrap samples are meant to be independent conditional on the original data. This means that, if the original data are generated from a set of random numbers $\boldsymbol{\omega}$, fresh random numbers are needed for each bootstrap sample. These random numbers are also drawings from the uniform distribution on Ω, independent of those used to generate the original data. With the notation of the previous section, then, a bootstrap statistic τ_j^*

is $\tau(b(\mu,\boldsymbol{\omega}),\boldsymbol{\omega}_j^*)$, where $\boldsymbol{\omega}_j^*$ is the set of random numbers used for the bootstrap sample, and $b(\mu,\boldsymbol{\omega})$ is the bootstrap DGP corresponding to the same data as those that gave the original test statistic $\hat{\tau} \equiv \tau(\mu,\boldsymbol{\omega})$.

We saw at the outset that rejection by the bootstrap test is the event $\tau < Q(\alpha,\mu^*)$. This event can, by (1), be written equivalently as $R(\tau,\mu^*) < \alpha$. In the notation we are now using, this event becomes $R(\tau(\mu,\boldsymbol{\omega}),b(\mu,\boldsymbol{\omega})) < \alpha$. The function R is a CDF, and so we can write down an explicit definition:

$$\begin{aligned} R(\alpha,\mu) &= \Pr{}_{\mu}(\tau < \alpha) = P\big(\{\boldsymbol{\omega} \,|\, \tau(\mu,\boldsymbol{\omega}) < \alpha\}\big) \\ &= \mathrm{E}\big(\mathrm{I}(\tau(\mu,\boldsymbol{\omega}) < \alpha)\big). \end{aligned} \tag{8}$$

Note that we need only one probability measure P in order to define this function, and that expectations can all be computed using this measure. The specific DGP appears in the definition of the random variable. For given arguments α and μ, the expectation (8) can be estimated by simulation in a direct fashion as

$$\frac{1}{N}\sum_{j=1}^{N} \mathrm{I}(\tau(\mu,\boldsymbol{\omega}_j) < \alpha), \tag{9}$$

where we draw N IID realisations of τ using the DGP μ.

If we generate B bootstrap samples for a bootstrap test, then, if the test statistic computed from the original data is $\tau(\mu,\boldsymbol{\omega})$, the bootstrap P value is

$$p_B^* = \frac{1}{B}\sum_{j=1}^{B} \mathrm{I}\big(\tau(b(\mu,\boldsymbol{\omega}),\boldsymbol{\omega}_j^*) < \tau(\mu,\boldsymbol{\omega})\big). \tag{10}$$

Note that, although in a simulation context, we know μ and ourselves generate $\boldsymbol{\omega}$, with real data we do not observe these things. However, since both τ and b are just deterministic functions of the data generated by the unobserved μ and $\boldsymbol{\omega}$, everything needed to evaluate (10) is observed.

As written, p_B^* is a random variable defined on a probability space that can be written as a $(B+1)$–fold product of the basic space $(\Omega,\mathcal{F},P)$. As $B \to \infty$, however, the strong law of large numbers applies, and p_B^* tends almost surely to

$$p^* \equiv \mathrm{E}\big(\mathrm{I}(\tau(b(\mu,\boldsymbol{\omega}),\boldsymbol{\omega}^*) < \tau(\mu,\boldsymbol{\omega})) \,|\, \mathcal{F}_{\boldsymbol{\omega}}\big) = R\big(\tau(\mu,\boldsymbol{\omega}),b(\mu,\boldsymbol{\omega})\big), \tag{11}$$

that is, to a random variable that is $\mathcal{F}_{\boldsymbol{\omega}}$-measurable, where $\mathcal{F}_{\boldsymbol{\omega}}$ is the sigma-algebra of the first factor in the $(B+1)$ fold product. We can thus define the function $p^*(\mu,\boldsymbol{\omega})$ as the right-hand side of (11).

The RP of the bootstrap test at level α is thus

$$\begin{aligned} \mathrm{E}\big(\mathrm{I}(p^*(\mu,\boldsymbol{\omega}) < \alpha)\big) &= \mathrm{E}\big(\mathrm{I}\big(R(\tau(\mu,\boldsymbol{\omega}),b(\mu,\boldsymbol{\omega})) < \alpha\big)\big) \\ &= \mathrm{E}\big(\mathrm{I}\big(\tau(\mu,\boldsymbol{\omega}) < Q(\alpha,b(\mu,\boldsymbol{\omega}))\big)\big) \\ &= \mathrm{E}\big(\mathrm{I}\big(R(\tau(\mu,\boldsymbol{\omega}),\mu) < R(Q(\alpha,b(\mu,\boldsymbol{\omega})),\mu)\big)\big) \\ &= \mathrm{E}\big(\mathrm{I}(p < r)\big), \end{aligned} \tag{12}$$

where the random variables p and r were defined in (2) and (3), and can now be written more explicitly as

$$p = R(\tau(\mu, \boldsymbol{\omega}), \mu) \quad \text{and} \quad r = R(Q(\alpha, b(\mu, \boldsymbol{\omega})), \mu).$$

Recall that p is distributed as $\mathrm{U}(0,1)$ for all μ, and that the CDF of r conditional on p under μ is denoted as $F(r \,|\, p)$.

4. Approximate Estimation of the Bootstrap Discrepancy

A fairly straightforward way to estimate the bootstrap RP, and hence the bootstrap discrepancy, is to use the approximations which, applied to a bootstrap distribution, yield the fast double bootstrap. These approximations are based on an assumption of the asymptotic independence of $\tau(\mu, \boldsymbol{\omega})$ and $b(\mu, \boldsymbol{\omega})$ for certain DGPs μ and certain sorts of bootstrap DGPs b. If there were complete independence, then it would imply the independence of p and q, and so the bootstrap RP (12) would become

$$\mathrm{E}\Big(\mathrm{E}\big(\mathrm{I}\big(\tau(\mu, \boldsymbol{\omega}) < Q(\alpha, b(\mu, \boldsymbol{\omega}))\big) \;\big|\; b(\mu, \boldsymbol{\omega})\big)\Big) = \mathrm{E}\big(R(Q(\alpha, b(\mu, \boldsymbol{\omega})), \mu)\big).$$

The next step in the approximation procedure is to consider the random variable sometimes denoted as τ^*. Realisations of this random variable under μ are generated by the formula

$$\tau^* = \tau(b(\mu, \boldsymbol{\omega}), \boldsymbol{\omega}^*).$$

We see that we need two sets of random numbers, $\boldsymbol{\omega}$ and $\boldsymbol{\omega}^*$, for each realisation of τ^*. The first set is used to generate a bootstrap DGP $b(\mu, \boldsymbol{\omega})$; the second set to generate a bootstrap statistic using this bootstrap DGP.

The distribution of τ^* under μ is calculated as follows.

$$\begin{aligned}
\Pr(\tau^* < \alpha) &= \mathrm{E}\big(\mathrm{I}(\tau(b(\mu, \boldsymbol{\omega}), \boldsymbol{\omega}^*) < \alpha)\big) \\
&= \mathrm{E}\Big(\mathrm{E}\big(\mathrm{I}(\tau(b(\mu, \boldsymbol{\omega}), \boldsymbol{\omega}^*) < \alpha) \;\big|\; \mathcal{F}_{\boldsymbol{\omega}}\big)\Big) \\
&= \mathrm{E}\big(R(\alpha, b(\mu, \boldsymbol{\omega}))\big) \equiv R^*(\alpha, \mu).
\end{aligned}$$

Thus the function $R^*(\alpha, \mu)$ is the CDF of τ^* under μ.

The quantile function Q^* for the distribution of τ^* is the inverse of R^*:

$$\alpha = R^*(Q^*(\alpha, \mu), \mu) = Q^*(R^*(\alpha, \mu), \mu)$$

for all $\alpha \in [0,1]$ and for all μ. Thus Q^* satisfies the relation

$$\mathrm{E}\big(R(Q^*(\alpha, \mu), b(\mu, \boldsymbol{\omega})\big) = \alpha.$$

This means that

$$\mathrm{E}\big(R(Q^*(\alpha, \mu), b(\mu, \boldsymbol{\omega})\big) = \mathrm{E}\big(R(Q(\alpha, b(\mu, \boldsymbol{\omega})), b(\mu, \boldsymbol{\omega})\big).$$

The approximation then consists of replacing the $b(\mu, \boldsymbol{\omega})$ where it appears as the second argument of the function R by μ, so that we have, approximately,

$$\mathrm{E}\big(R(Q(\alpha, b(\mu, \boldsymbol{\omega})), \mu\big) = R(Q^*(\alpha, \mu), \mu).$$

A simulation estimate of the right-hand side of this relation, easily obtained by the procedure used for the FDB, is therefore an *approximate* simulation estimate of the bootstrap RP.

5. A Better Approximation

If $\tau(\mu, \boldsymbol{\omega})$ and $b(\mu, \boldsymbol{\omega})$ are independent, then, as we saw at the beginning, the bootstrap discrepancy is the expectation of the random variable q, and the bootstrap RP is $\alpha + \mathrm{E}(q)$, which is just

$$\mathrm{E}\big(R(Q(\alpha, b(\mu, \boldsymbol{\omega})), \mu)\big). \tag{13}$$

A simulation estimate of this expectation is then approximate only on account of simulation randomness and the approximation inherent in the independence assumption.

The function $R(\cdot, \mu)$ is easily estimated by simulation, as in (9). The procedure for estimating the expectation (13) is a little more complicated. It works as follows. First draw $N+1$ sets of random numbers, $\boldsymbol{\omega}_j$, $j = 1, \ldots, N+1$. Use these to generate $N+1$ realisations $\tau(\mu, \boldsymbol{\omega}_j)$ and $b(\mu, \boldsymbol{\omega}_j)$, and store the results. For a parametric bootstrap, each $b(\mu, \boldsymbol{\omega}_j)$ is represented by a set of parameter estimates, and for a nonparametric bootstrap by a set of objects to be resampled, possibly along with a set of parameter estimates.

Now comes a loop in which we obtain $N+1$ realisations of the α-quantile of the bootstrap distribution. In the iteration of the loop indexed by j, an inner loop is required, in which N samples are drawn from the $b(\mu, \boldsymbol{\omega}_j)$ distribution, using the random numbers $\boldsymbol{\omega}_k$, $k \neq j$, and statistics τ_{jk} computed for each of them. The α-quantile of the distribution of τ under $b(\mu, \boldsymbol{\omega}_j)$ can be estimated as the $\lfloor \alpha N \rfloor + 1$ order statistic of the τ_{jk}. Now we can count the number of the $\tau(\mu, \boldsymbol{\omega}_k)$, $k = 1, \ldots, N+1$, $k \neq j$, which are more extreme than the estimated quantile. This number, divided by N, is then an estimate of $R(Q(\alpha, b(\mu, \boldsymbol{\omega}_j)), \mu)$. Finally, we average the $N+1$ estimates thus obtained, and obtain an estimate of (13).

A concrete example should aid comprehension of the procedure. Consider the simple model

$$y_t = \beta + \sigma u_t, \quad u_t \sim \mathrm{IID}(0, 1), \quad s, t = 1, \ldots, n. \tag{14}$$

we wish to study a bootstrap test of the hypothesis that $\beta = 0$. The obvious test statistic is a t statistic, defined as

$$t(\boldsymbol{y}) = \frac{n^{-1} \sum_{t=1}^{n} y_t}{\hat{\sigma}/n^{1/2}} \tag{15}$$

with $\hat{\sigma}^2 = (n-1)^{-1} \sum_{t=1}^{n} (y_t^2 - \bar{y})^2$, $\bar{y} = n^{-1} \sum_{t=1}^{n} y_t$. If the data y_t are generated from $(\mu, \boldsymbol{\omega})$, then we have

$$\tau(\mu, \boldsymbol{\omega}) = t(\boldsymbol{y}(\mu, \boldsymbol{\omega})),$$

using the mapping $\boldsymbol{y}(\mu, \boldsymbol{\omega})$ defined at the end of section 2.

The bootstrap DGP we use is a nonparametric one, with resampling of rescaled residuals from the estimation of β. Specifically,

$$y_t^* = u_t^*, \quad u_t^* \sim \mathrm{IID}\big(\mathrm{EDF}((n/(n-1))^{1/2} \hat{u}_t)\big)$$

in conventional notation. Thus the bootstrap DGP $b(\mu, \boldsymbol{\omega})$ is completely characterised by the $n-$vector

$$\boldsymbol{y}(\mu, \boldsymbol{\omega}) - (n^{-1} \boldsymbol{\iota}^\top \boldsymbol{y}(\mu, \boldsymbol{\omega})) \boldsymbol{\iota},$$

where $\boldsymbol{\iota}$ is an $n-$vector of ones. Denote this vector by $\boldsymbol{u}(\mu, \boldsymbol{\omega})$, with typical element $u_t(\mu, \boldsymbol{\omega})$.

The DGP μ must specify the distribution of the random elements u_t in (14). Let this distribution, which should have first and second moments equal to 0 and 1 respectively, have CDF F. The vectors $\boldsymbol{\omega}$ of random numbers are $n-$vectors of independent drawings from U(0, 1). Then we have

$$y_t(\mu, \boldsymbol{\omega}) = \sigma F^{-1}(\omega_t), \tag{16}$$

assuming that μ satisfies the null hypothesis that $\beta = 0$. The nuisance parameter σ can be chosen arbitrarily, since the statistic is scale independent. In more general cases, it would be a parameter of the simulation experiment.

For a bootstrap sample generated using the DGP $b(\mu, \boldsymbol{\omega})$ by the random numbers $\boldsymbol{\omega}^*$, we have

$$y_t(b(\mu, \boldsymbol{\omega}), \boldsymbol{\omega}^*) = \text{Order statistic } \lfloor n\omega_t^* \rfloor + 1 \text{ of the } u_t(\mu, \boldsymbol{\omega}). \tag{17}$$

In this way, we define $y_t(b(\mu, \boldsymbol{\omega}), \boldsymbol{\omega}^*)$ as the ω_t^* quantile of the EDF of the $u_t(\mu, \boldsymbol{\omega})$. It is thus convenient to sort the $u_t(\mu, \boldsymbol{\omega})$ in increasing order before storing them, in order to be able to extract order statistics easily.

The simulation experiment can now be described as follows.

- For each $i = 1, \ldots, N + 1$, draw the sample $\boldsymbol{y}(\mu, \boldsymbol{\omega}_i)$ using the formula (16), and use it to compute the test statistic $\tau(\mu, \boldsymbol{\omega}_i)$ and the vector $\boldsymbol{u}(\mu, \boldsymbol{\omega}_i)$ of sorted rescaled residuals.
- For each $i = 1, \ldots, N+1$, draw N simulated samples $\boldsymbol{y}(b(\mu, \boldsymbol{\omega}_i), \boldsymbol{\omega}_j)$, $j = 1, \ldots, N+1$, $j \neq i$, using the formula (17). Compute the N statistics

  $$\tau_{ij} \equiv t(\boldsymbol{y}(b(\mu, \boldsymbol{\omega}_i), \boldsymbol{\omega}_j) \tag{18}$$

  using (15). Obtain an estimate of the α-quantile of the distribution of the τ_{ij} as the $\lfloor \alpha N \rfloor + 1$ order statistic, q_i, of the τ_{ij}. Compute the frequency f_i as

  $$f_i = \frac{1}{N} \#\{j, j = 1, \ldots, n + 1, j \neq i \,|\, \tau(\mu, \boldsymbol{\omega}_j) < q_i\}. \tag{19}$$

- Estimate (13) as the average of the f_i, $i = 1, \ldots, N + 1$.

Depending on the specific nature of the problem at hand, it may be easier and faster to compute the τ_{ij} by fixing j and computing the statistics for all $i \neq j$, rather than by fixing i and computing them for all $j \neq i$. For the J test, computing time was reduced from a little over an hour to one minute and thirty seconds by this device, which made it possible to vectorise the calculations much more efficiently.

The N statistics τ_{ij} of (18) are, for given i, IID conditional on $\mathcal{F}_{\boldsymbol{\omega}_i}$. Of course, if we allowed $j = i$, then τ_{ii} would be a drawing from a different distribution from that of the τ_{ij}, $j \neq i$. By excluding τ_{ii} from the set (18), we thus obtain in q_i a good estimate of $Q(\alpha, b(\mu, \boldsymbol{\omega}_i))$.

In the limit as $N \to \infty$, the q_i are thus also IID, since the $Q(\alpha, b(\mu, \boldsymbol{\omega}_i))$ are IID. Similarly, in the limit as $N \to \infty$, the f_i are IID, since they are estimates of the $R(Q(\alpha, b(\mu, \boldsymbol{\omega}_i), \mu)$.

Once again, we must exclude $\tau(\mu, \boldsymbol{\omega}_i)$ from the set of realisations used to compute f_i, since it is not independent of q_i. With this precaution, we obtain a consistent and asymptotically normal estimate of (13).

The realisations $\tau(\mu, \boldsymbol{\omega}_j)$ used in (19) are *not* independent of the $t(\boldsymbol{y}(b(\mu, \boldsymbol{\omega}_i), \boldsymbol{\omega}_j)$ of (18). On the contrary, for each i, $t(\boldsymbol{y}(b(\mu, \boldsymbol{\omega}_i), \boldsymbol{\omega}_j)$ is strongly correlated with $\tau(\mu, \boldsymbol{\omega}_j)$, since the same random numbers $\boldsymbol{\omega}_j$ are used to generate the two statistics. This is highly desirable, since the correlation leads to a reduction in the variance of q_i, as seen in the Appendix.

6. The Best Approximation

Recall that the conditional expectation in (6) is a very accurate approximation to the bootstrap discrepancy in all cases so far considered. Thus the bootstrap RP is very close to $\alpha + \mathrm{E}(q \,|\, \alpha)$, which, from (3), can be written as $\mathrm{E}(R(Q(\alpha, \mu^*), \mu) \,|\, p = \alpha)$. In our more explicit notation, this becomes

$$\mathrm{E}\big(R(Q(\alpha, b(\mu, \boldsymbol{\omega})), \mu) \,|\, R(\tau(\mu, \boldsymbol{\omega}), \mu) = \alpha\big). \tag{20}$$

In the previous section, we saw how to simulate the variable of which we want to evaluate the conditional expectation, the f_i being estimates of the $R(Q(\alpha, b(\mu, \boldsymbol{\omega}_i)), \mu)$. Now, we wish to be able to simulate simultaneously the realisation $R(\tau(\mu, \boldsymbol{\omega}_i), \mu)$. The $\tau(\mu, \boldsymbol{\omega}_i)$ are computed in the course of the simulation experiment of the previous section, and, altogether, they provide an estimate of the CDF R *via* the EDF of the $\tau(\mu, \boldsymbol{\omega}_i)$. Thus $R(\tau(\mu, \boldsymbol{\omega}_i), \mu)$ is estimated as the rank of $\tau(\mu, \boldsymbol{\omega}_i)$ in the set of the $\tau(\mu, \boldsymbol{\omega}_j)$, $j \neq i$, computed as the number of the $\tau(\mu, \boldsymbol{\omega}_j)$ that are less than $\tau(\mu, \boldsymbol{\omega}_i)$, divided by N. The conditional expectation (20) is thus the expectation of $R(Q(\alpha, b(\mu, \boldsymbol{\omega})), \mu)$ conditional on the rank of $\tau(\mu, \boldsymbol{\omega})$.

There are various ways to estimate (20), but the most straightforward is to use a kernel estimate. The first step is to sort the $R(Q(\alpha, b(\mu, \boldsymbol{\omega}_i)), \mu)$ in increasing order of the $\tau(\mu, \boldsymbol{\omega}_i)$. Denote the sorted realisations by r_i, $i = 0, 1, \ldots, N$. The value of $R(\tau(\mu, \boldsymbol{\omega}_i), \mu)$ for realisation i in the sorted array is then i/N. If we use a kernel k and a bandwidth h, the Nadaraya-Watson estimate of the conditional expectation is

$$\sum_{i=0}^{N} k\Big(\frac{i/N - \alpha}{h}\Big)\, r_i \;\Big/\; \sum_{i=0}^{N} k\Big(\frac{i/N - \alpha}{h}\Big).$$

Another, probably better, estimator is the locally linear estimator of the conditional expectation, which is found by solving the estimating equation

$$\sum_{i=0}^{N} k\Big(\frac{i/N - \alpha}{h}\Big)\big(r_i - m(\alpha) - m'(\alpha)(\frac{i}{N} - \alpha)\big) = 0$$

for the estimate $m(\alpha)$.

Of course, however well the expression $\alpha + \mathrm{E}(q \,|\, \alpha)$ approximates the bootstrap RP, it remains an approximation to the true expression (4), or the expression (6) for the bootstrap discrepancy plus α. It is possible to estimate (4) directly, again using kernel estimators. We begin by seeing how to estimate the conditional CDF $F(r \,|\, p)$. The EDF of a paired IID sample of $N+1$ realisations (p_i, r_i), $i = 0, 1, \ldots, N$, is

$$F(p, r) = \frac{1}{N+1} \sum_{i=0}^{N} \mathrm{I}(p_i < p)\mathrm{I}(r_i < r),$$

where, with our sorted sample, $p_i = i/N$ and $r_i = R(Q(\alpha, b(\mu, \boldsymbol{\omega}_i), \mu), \mu)$. Replacing the indicator functions by a cumulative kernel K, and using bandwidth h, we get the usual kernel estimator of this joint CDF:

$$\hat{F}(p, r) = \frac{1}{N+1} \sum_{i=0}^{N} K\Big(\frac{r - r_i}{h}\Big) K\Big(\frac{p - p_i}{h}\Big).$$

The joint density is thus estimated as

$$\hat{f}(p, r) = \frac{1}{(N+1)h^2} \sum_{i=0}^{N} k\Big(\frac{r - r_i}{h}\Big) k\Big(\frac{p - p_i}{h}\Big),$$

where $k = K'$ is the kernel that corresponds to the cumulative kernel K. The conditional density is then estimated by the joint density estimate divided by the kernel estimate of the marginal density of p:

$$\hat{f}(r \,|\, p) = \frac{1}{(N+1)h^2} \sum_{i=0}^{N} k\Big(\frac{r - r_i}{h}\Big) k\Big(\frac{p - p_i}{h}\Big) \Big/ \frac{1}{(N+1)h} \sum_{i=0}^{N} k\Big(\frac{p - p_i}{h}\Big).$$

By integrating with respect to r, we obtain an estimate of the CDF of r conditional on p:

$$\hat{F}(r \,|\, p) = \sum_{i=0}^{N} K\Big(\frac{r - r_i}{h}\Big) k\Big(\frac{p - p_i}{h}\Big) \Big/ \sum_{i=0}^{N} k\Big(\frac{p - p_i}{h}\Big).$$

We now see that there is in fact no need to use the cumulative kernel, since it can be replaced by an indicator function, thereby eliminating bias at the cost of slight non-smoothness. Since we intend to integrate the function, this is of no consequence. What we actually use is thus

$$\hat{F}(r \,|\, p) = \sum_{i=0}^{N} \mathrm{I}(r_i < r)\, k\Big(\frac{p - p_i}{h}\Big) \Big/ \sum_{i=0}^{N} k\Big(\frac{p - p_i}{h}\Big).$$

Finally, then, the true bootstrap RP can be estimated, using (4), by

$$1 - \int_0^1 \hat{F}(p \,|\, p)\, dp. \tag{21}$$

A proper numerical integration algorithm, such as Romberg integration, seems necessary for the integral to be estimated with sufficient accuracy. A Riemann sum defined on a grid of equally spaced points separated by 0.01 was *not* accurate enough.

Of course, there is a way to simulate the bootstrap RP that is much simpler than all the methods discussed so far. Since we have a paired IID sample of realisations (p_i, r_i), and since rejection by the bootstrap test is the event $p < r$, we can estimate the bootstrap RP directly as the proportion

$$\frac{1}{N+1}\sum_{i=0}^{N} \mathrm{I}(p_i < r_i). \tag{22}$$

This does indeed give results not very different from those given by (21). But this approach does not impose the fact that the marginal distribution of the random variable p is $\mathrm{U}(0,1)$. Thus we can expect that (21) will benefit from a simulation efficiency gain relative to (22).

Simpler yet is to construct the bootstrap P value for each set of realisations. According to (11), the bootstrap P value is $R\big(\tau(\mu,\boldsymbol{\omega}), b(\mu,\boldsymbol{\omega})\big)$. For realisation i, we substitute $\boldsymbol{\omega}_i$ for $\boldsymbol{\omega}$ in this expression, and estimate the value of the expression by

$$p_i^* \equiv \frac{1}{N}\sum_{j=1}^{N} \mathrm{I}\big(\tau(b(\mu,\boldsymbol{\omega}_i),\boldsymbol{\omega}_j^*) < \tau(\mu,\boldsymbol{\omega}_i)\big).$$

The bootstrap RP at level α is then estimated as the proportion of the p_i^* smaller than α. These two simpler approaches give very similar results.

Appendix

Suppose that we have two CDFs, F and G, and we wish to estimate the quantity $F(G^{-1}(\alpha))$, for some given $\alpha \in [0,1]$. A sample of pairs (x_i, y_i), $i = 1, \ldots, N$, is available, drawn from a joint distribution with the property that the marginal distribution of the x_i has CDF F, and that of the y_i has CDF G. The obvious estimator is then $\hat{F}(\hat{G}^{-1}(\alpha))$, where $\hat{F}$ and $\hat{G}$ are the EDFs of the x_i and the y_i respectively. We choose the number N so that αN is an integer, which means that $\hat{G}^{-1}(\alpha)$ is a well-defined order statistic of the y_i.

We use the delta method to estimate the variance of the estimate $\hat{F}(\hat{G}^{-1}(\alpha))$. To that end, write $\hat{G} = G + \delta G$ and $\hat{G}^{-1} = G^{-1} + \delta G^{-1}$. Then, since $\hat{G}(\hat{G}^{-1}(\alpha)) = \alpha$, we have, to first order in small quantities,

$$\alpha = (G + \delta G)(G^{-1}(\alpha)) + G'(G^{-1}(\alpha))(\delta G^{-1})(\alpha).$$

Thus, working throughout only to first order in small quantities,

$$(\delta G^{-1})(\alpha) = -\frac{\delta G(G^{-1}(\alpha))}{G'(G^{-1}(\alpha))}.$$

Consequently, we see that, letting $\hat{F} = F + \delta F$,

$$\begin{aligned}\hat{F}(\hat{G}^{-1}(\alpha)) &= (F + \delta F)(G^{-1} + (\delta G^{-1}))(\alpha) \\ &= F(G^{-1}(\alpha)) + \delta F(G^{-1}(\alpha)) - \frac{F'(G^{-1}(\alpha))}{G'(G^{-1}(\alpha))}\delta G(G^{-1}(\alpha)),\end{aligned}$$

so that

$$N^{1/2}\big(\hat{F}(\hat{G}^{-1}(\alpha)) - F(G^{-1}(\alpha))\big) = \begin{bmatrix} 1 & -\dfrac{F'(G^{-1}(\alpha))}{G'(G^{-1}(\alpha))} \end{bmatrix} \begin{bmatrix} N^{1/2}\delta F(G^{-1}(\alpha)) \\ N^{1/2}\delta G(G^{-1}(\alpha)) \end{bmatrix}.$$

For the asymptotic variance of the estimator $\hat{F}(\hat{G}^{-1}(\alpha))$, we require that of the 2-vector that is the second factor on the right-hand side above. As usual, we have

$$\lim_{n\to\infty} \mathrm{Var}(N^{1/2}\delta F(x)) = F(x)(1 - F(x)) \text{ and } \lim_{n\to\infty} \mathrm{Var}(N^{1/2}\delta G(x)) = G(x)(1 - G(x)).$$

For the covariance, note that

$$\begin{aligned}\mathrm{E}\big(\hat{F}(x)\hat{G}(y)\big) &= N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{E}\big(\mathrm{I}(x_i < x)\mathrm{I}(y_j < y)\big) \\ &= N^{-2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}\mathrm{E}\big(\mathrm{I}(x_i < x)\mathrm{I}(y_j < y)\big) + N^{-2}\sum_{i=1}^{N}\mathrm{E}\big(\mathrm{I}(x_i < x)\mathrm{I}(y_i < y)\big) \\ &= (N-1)F(x)G(y)/N + H(x,y)/N,\end{aligned}$$

where H is the *joint* CDF of the (x_i, y_i). Therefore

$$\begin{aligned}\mathrm{cov}(N^{1/2}\delta F(x), N^{1/2}\delta G(x)) &= N\,\mathrm{cov}(\hat{F}(x), \hat{G}(x)) \\ &= N\mathrm{E}(\hat{F}(x)\hat{G}(y)) - N\mathrm{E}(\hat{F}(x))\mathrm{E}(\hat{G}(y)) \\ &= (N-1)F(x)G(y) + H(x,y) - NF(x)G(y) \\ &= H(x,x) - F(x)G(x).\end{aligned}$$

We conclude that

$$\begin{aligned}\lim_{n\to\infty} \mathrm{Var}\Big(N^{1/2}\big(\hat{F}(\hat{G}^{-1}(\alpha)) - F(G^{-1}(\alpha))\big)\Big) &= \\ \begin{bmatrix} 1 & -\dfrac{F'}{G'} \end{bmatrix} \begin{bmatrix} F(1-F) & H - FG \\ H - FG & G(1-G) \end{bmatrix} \begin{bmatrix} 1 \\ -\dfrac{F'}{G'} \end{bmatrix} & \\ = F(1-F) + \left(\frac{F'}{G'}\right)^2 G(1-G) - 2\frac{F'}{G'}(H - FG), & \qquad (23)\end{aligned}$$

where all the functions in the expression of the variance are evaluated at $G^{-1}(\alpha)$. Since $G(G^{-1}(\alpha)) = \alpha$, we may simplify further to obtain the following expression for the asymptotic variance:

$$F(1-F) + \left(\frac{F'}{G'}\right)^2 \alpha(1-\alpha) - 2\frac{F'}{G'}(H - \alpha F).$$

If the x_i and y_i are positively correlated, then $H(\alpha, \alpha) > F(\alpha)G(\alpha)$. Since F' and G' are nonnegative, this implies that the variance of $\hat{F}(\hat{G}(\alpha))$ is smaller in the presence of positive correlation than if the x_i and y_i were uncorrelated. Obviously everything in (23) can be estimated consistently, the derivatives F' and G' by kernel methods. If we have approximately that $F(\alpha) = G(\alpha) = \alpha$, so that $F'(G^{-1}(\alpha)) = G'(F^{-1}(\alpha)) = 1$, then the variance is approximately

$$2\big(\alpha(1-\alpha) - (H(\alpha,\alpha) - \alpha^2)\big). \tag{24}$$

If there were perfect correlation, then we have $x_i = y_i$, so that $H(\alpha, \alpha) = \alpha$. In this case, as expected, the variance (24) is zero.